Biochem. J. (1985) **229**, 189–196
Printed in Great Britain

Primary structure of the telopeptide and a portion of the helical domain of chicken type II procollagen as determined by DNA sequence analysis

Ferenc DEÁK, W. Scott ARGRAVES, Ibolya KISS, Kenneth J. SPARKS and Paul F. GOETINCK*
Department of Animal Genetics and Molecular Genetics and Cell Biology Section, University of Connecticut, Storrs, CT 06268, U.S.A.

(*Received 6 November 1984/21 March 1985; accepted 25 March 1985*)

A comparison of the nucleotide sequences of three new cDNA clones for chicken type II procollagen with the sequences of the other three types of chicken fibrillar procollagens reveals that the most conserved regions correlate with the positions of hydroxyproline, hydroxylysine, cysteine and lysine residues. On the basis of replacement-site-divergence calculations it is concluded that α1(II) and α1(I) procollagens diverged later than α1(I) and α2(I) procollagens.

Collagens represent a family of extracellular matrix proteins that contain at least one triple-helical domain and form supramolecular aggregates either alone or in conjunction with other extracellular matrix components (Ninomiya *et al.*, 1984*a*). Although the collagens are very similar in structure, they show structural and functional diversity in connection with their characteristic tissue distribution (Fessler & Fessler, 1978; Bornstein & Sage, 1980).

An extensive sequence comparison is required in order to understand the differences in the structure and function of different collagens as well as the evolution of these ancient genes. The acquisition of recombinant DNA clones coding for collagens permits structural analysis at both the nucleotide and the amino acid level. Amino acid and nucleotide sequences of the C-propeptide, telopeptide and part of the helical domain have been published for chicken (Fuller & Boedtker, 1981), human (Bernard *et al.*, 1983*a*,*b*), mouse (Monson *et al.*, 1982) and sheep (Boyd *et al.*, 1980), type I and chicken type III procollagens (Yamada *et al.*, 1983).

Further, chicken (Vuorio *et al.*, 1982; Lukens *et al.*, 1983; Sandell *et al.*, 1983; Ninomiya *et al.*, 1984*b*) and human (Strom & Upholt, 1984) type II procollagen clones have been isolated and nucleotide sequences of the C-propeptide and part of the telopeptide have been reported. While this manuscript was in preparation, Sandell *et al.* (1984) published the nucleotide sequence of the 3′-end of the chicken type II procollagen gene encoding for the C-propeptide, the C-telopeptide and the last 15 amino acids of the triple-helical domain.

In the present paper we report the first case of the isolation and characterization of cDNA clones containing the coding sequences for the entire C-propeptide and the C-telopeptide as well as 96 amino acid residues of the helical domain of the chicken type II procollagen chain. A thorough sequence comparison with evolutionary implications is also presented. A preliminary report has been published elsewhere (Deák *et al.*, 1984).

Abbreviations used: poly(A)$^+$ RNA, polyadenylated RNA; bp, base-pair; kb, kilobase; SSC, 0.15M-NaCl/0.015M-sodium citrate buffer, pH 7.0.

* To whom correspondence and requests for reprints should be sent.

Experimental

Materials

α-^{32}P-labelled deoxynucleotide triphosphates and [γ-^{32}P]ATP were supplied from New England Nuclear. T4-bacteriophage polynucleotide kinase and restriction endonucleases were purchased from New England Biolabs. Nitrocellulose filters were from Schleicher and Schuell. Sea-Kem ME agarose was from FMC Corporation. Piperidine, formic acid, phenol and urea were from Fisher Scientific Co. Dimethyl sulphate was obtained from Aldrich Chemical Co. Hydrazine was from Eastman Organic Chemicals. Acrylamide and bisacrylamide were purchased from Bio-Rad Laboratories.

Construction and screening of a chicken sternal-cartilage cDNA library

RNA preparation, synthesis and cloning of cDNA are not described in detail in the present

paper. Briefly, cDNA was synthesized by using avian-myeloblastosis-virus reverse transcriptase, oligo(dT) primer and poly(A)$^+$ RNA isolated from chicken-embryo sterna. The double-stranded cDNA was inserted into pUC8 and pUC9 vectors after sequential addition of *Sal*I and *Eco*RI linkers, as described by Helfman *et al.* (1983). Transformation of *Escherichia coli* DH1 was done as described by Hanahan (1983). Filter-bound colonies were screened by following the procedure of Hanahan & Meselson (1983), with labelled poly(A)$^+$ RNA isolated from calvaria and sternal cartilage of 14-day-old chick embryos. The RNA was alkali-fragmented and end-labelled by using [γ-^{32}P]ATP and polynucleotide kinase by the method of Ninomiya *et al.* (1984*b*). Those clones hybridizing most strongly to sternal RNA but not to calvarial RNA were isolated for further analysis.

Northern analysis and dot hybridization

Poly(A)$^+$ RNA was electrophoresed on 2.2M-formaldehyde/0.8% agarose gels (20mM-sodium acetate), transferred to nitrocellulose filters (SS BA 85) and hybridized to 20ng (1×10^7 d.p.m.) of nick-translated DNA in solution containing 50% formamide, 2×SSC, 1×Denhardt's solution, 0.1% sodium dodecyl sulphate and 50μg of denatured salmon sperm DNA/ml at 42°C for 24h after pre-hybridizing the filters under the same conditions without the probe for 16h. The filters were washed at room temperature with solution containing 2×SSC and 0.1% sodium dodecyl sulphate, then at 68°C with solution containing 2×SSC and 0.1% sodium dodecyl sulphate, followed by solution containing 1×SSC and 0.1% sodium dodecyl sulphate and solution containing 0.2×SSC and 0.1% sodium dodecyl sulphate for 10min each before autoradiography (Thomas, 1980).

RNA dot blotting was done with a Schleicher and Schuell Minifold apparatus. RNA was mixed in 100μl of 10×SSC buffer containing 2.5M-formaldehyde, heated for 5min at 68°C and then chilled on ice. Samples were applied to the filters under vacuum and each well was rinsed with 200μl of the buffer. The filters were hybridized in accordance with conditions outlined for the Northern analysis.

Nucleotide sequencing and sequence comparison

Cleared bacterial lysates were made by using the procedure of Clewell & Helinski (1972). Plasmid DNA was purified by using CsCl/ethidium bromide-gradient centrifugation (Tanaka & Weisblum, 1975). All restriction-endonuclease digestions were done in accordance with the suggestions of Fuchs & Blakesley (1983).

DNA sequencing was done by the chemical method of Maxam & Gilbert (1980). DNA fragments were 5'-end-labelled with ^{32}P by the exchange reaction with the use of T4 polynucleotide kinase (Berkner & Folk, 1977). End-labelled fragments were either asymmetrically cleaved with a second restriction endonuclease or strand-separated and then purified by electroelution from polyacrylamide gels. Chemically cleaved DNA fragments were electrophoresed on 8% and 20% polyacrylamide sequencing gels as well as 6%-buffer-gradient sequencing gels (Biggin *et al.*, 1983). Dupont Cronex film was used with intensifying screens for autoradiography. Analysis of DNA sequences was performed with the aid of an Apple II computer and software developed in this laboratory.

The divergence of nucleotide sequences was examined by a modification of the method of Perler *et al.* (1980) in which the reliability of the statistical comparisons was improved by calculating variance-weighted average divergence, as suggested by Bernard *et al.* (1983*b*).

Results and discussion

Screening of chicken sternal-cartilage cDNA library

The cDNA library was screened for clones coding for abundant sternal-cartilage mRNA species. About 6000 filter-bound colonies were hybridized to ^{32}P-labelled sternal poly(A)$^+$ RNA (Hanahan & Meselson, 1983). Plasmid DNA was isolated from 100 colonies that gave strong signals in the colony hybridization test. Inserts from a set of these plasmids were ^{32}P-labelled and hybridized to sternal and calvarial poly(A)$^+$ RNA in a Northern RNA-transfer experiment (Thomas, 1980). As shown in Fig. 1(*a*), the insert isolated from one clone hybridized strongly to a sternal-cartilage RNA species slightly larger than 27S. This transcript is faintly apparent in poly(A)$^+$ RNA isolated from chicken-embryo calvaria. RNA extracted from embryonic limbs was used in a dot hybridization experiment to demonstrate that the RNA hybridizing to the clone appears concomitant with limb development (Fig. 1*b*).

Identification of clones for chicken type II procollagen

On the basis of the determined transcript size and tissue specificity, we pursued the possibility that the clone, below called pCgII-S01, coded for chicken type II procollagen. First, a detailed restriction map of the insert was constructed with the use of 15 different restriction enzymes (Fig. 2*a*). This map was compared with the restriction map for type II procollagen cDNA published by Ninomiya *et al.* (1984*b*). Comparison of overlapping regions showed basically identical restriction patterns except that pCgII-S01 contained a *Bgl*I-cleavage and some *Bst*NI-cleavage sites not pre-

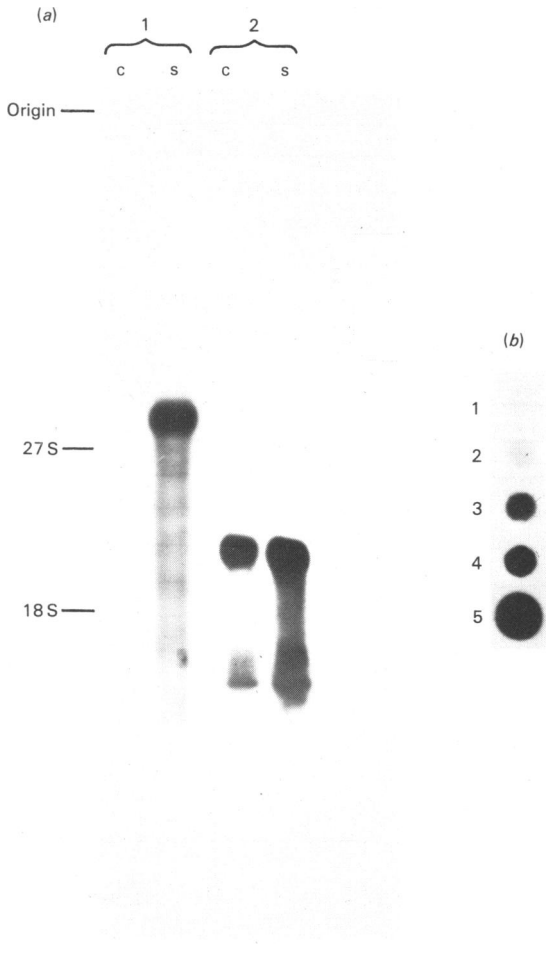

Fig. 1. *Hybridization of sternal-cartilage RNA-positive clones to filter-bound RNA*

(*a*) Size determination of RNA complementary to pCgII-S01 (1) and pP03F08 (2), a clone hybridizing to an abundant mRNA species present in both sternal and calvarial tissues. A 2μg portion of poly(A)$^+$ RNA isolated from embryonic sternum (s) or calvaria (c) was applied per lane, electrophoresed, transferred and hybridized as described in the Experimental section. (*b*) Developmental appearance of RNA complementary to PCgII-S01 during limb development. Portions (4.5μg) of total limb RNA isolated from 4-, 5-, 6- and 7-day-old embryos (dots 1, 2, 3 and 4 respectively) were bound to a nitrocellulose filter and hybridized to ^{32}P-labelled pCgII-S01 insert. Hybridization of 3μg of sternal poly(A)$^+$ RNA (dot 5) is shown for comparison.

sent in the published sequence (Ninomiya *et al.*, 1984*b*). The explanation for these differences is not known; however, they can be derived from the known sequence by assuming single base-pair mutations. They may represent DNA polymorphism due to different genetic backgrounds.

Restriction mapping also revealed the presence of many *Sau*96I-cleavage and *Nci*I-cleavage sites at the 5′-end of the pCgII-S01 insert. This indicated the frequent occurrence of sequences potentially coding for Gly-Pro and Pro-Gly residues and therefore the presence of collagen helical-domain coding regions. The presence of Gly-Xaa-Yaa repeats was confirmed by nucleotide sequence analysis (see below). These observations led us to conclude that PCgII-S01 coded for the 3′-terminal portion of chicken type II procollagen mRNA.

Nucleotide sequence analysis

Restriction mapping revealed that pCgII-S01 contained a 1.54kb insert that encoded the 3′-terminal untranslated region, the complete C-propeptide, the telopeptide and part of the helical domain. This is the largest type II procollagen cDNA clone isolated to date. Using ^{32}P-labelled total insert and the 255bp *Taq*I fragment from pCgII-S01, we screened the library for more clones for type II procollagen. Two of the positive clones, pCgII-S05 and pCgII-S13, were further analysed by detailed restriction-endonuclease mapping. Fig. 2(*b*) shows the composite restriction map of these clones, indicating their relative location with respect to the polypeptide (Fig. 2*c*). pCgII-S05 has a 1460bp insert coding for about 75bp more of the 3′-terminal untranslated region than pCgII-S01, and pCgII-S13 contains a 900bp insert that extends into the helical domain 50bp further than pCgII-S01. The formation of pCgII-S13 is not understandable unless we assume that it was primed at the relatively A-rich region preceding the *Bam*HI-cleavage site.

Nucleotide sequences reported here were derived from pCgII-S01 and pCgII-S13. The sequencing strategy is summarized in Fig. 2(*b*).

In addition to the helical-domain-and-telopeptide-coding region, a portion of the C-propeptide-coding region of pCgII-S01 was sequenced in order to confirm that the insert coded for type II procollagen. The nucleotide sequence of this region was essentially identical with that of the type II procollagen published by Ninomiya *et al.* (1984*b*). (Out of 340 base-pairs sequenced, only three base-pair differences were noted.) In positions 85–90 we found CGC AGC instead of the reported CGA GCC, and in position 522 we found G (sequence not published here) instead of A (Ninomiya *et al.*, 1984*b*).

Both DNA strands were sequenced in the helical-domain-and-telopeptide-coding region of pCgII-S01. Additional helical-domain-coding sequences were also obtained from pCgII-S13. The composite nucleotide and the deduced amino acid sequences of the *C*-terminal parts of the helical region (288bp), the entire telopeptide (81bp) and a

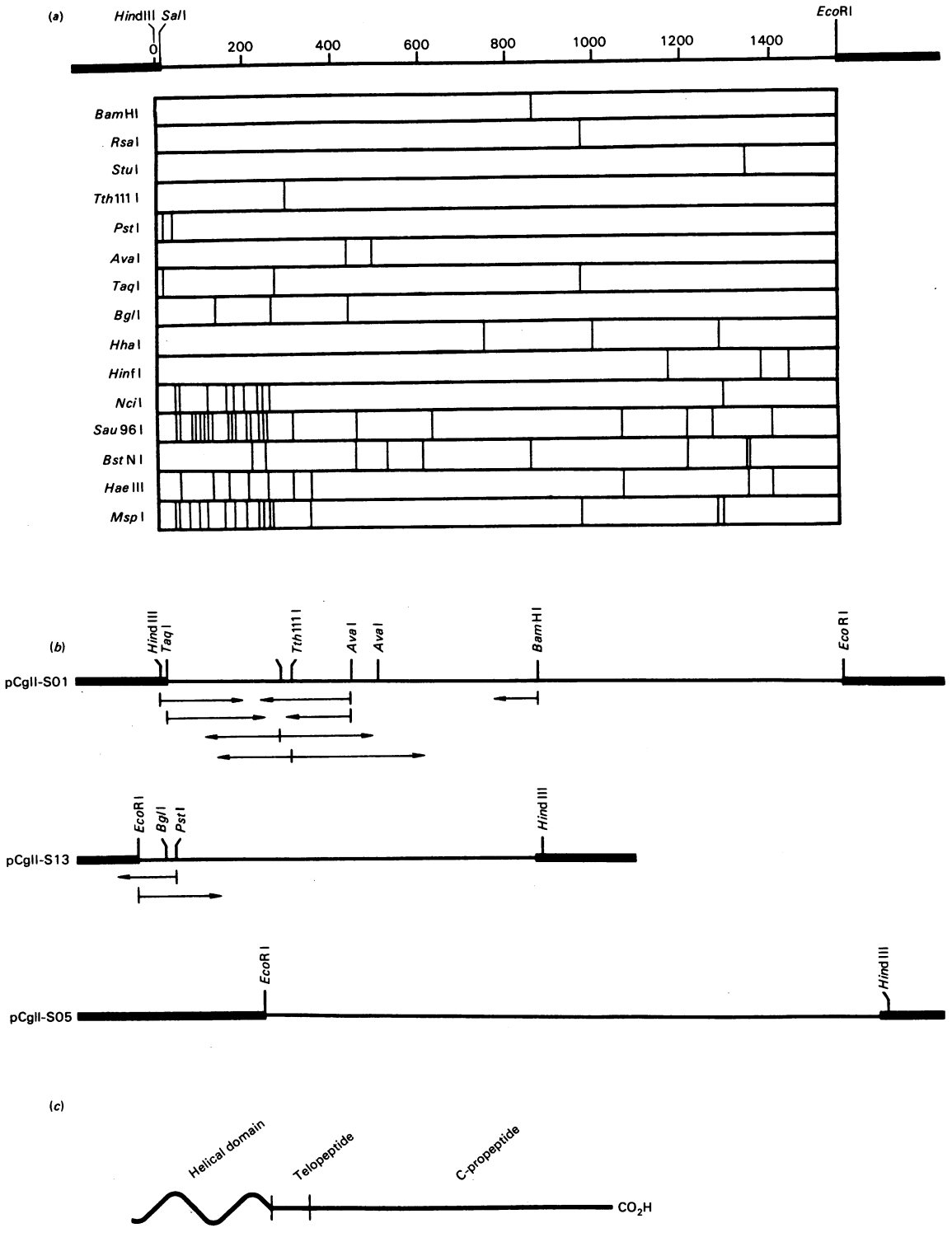

Fig. 2. *Composite restriction-endonuclease map of the clones for α1(II) procollagen*
(*a*) Restriction-endonuclease cleavage map of pCgII-S01. (*b*) Sequencing strategy for cDNA clones for type II procollagen. (*c*) Schematic representation of the domain structure of type II procollagen in alignment with the insert.

portion of the C-propeptide are shown in Fig. 3. The nucleotide sequence contained many 6–10 bp-long tandem repeats between positions −292 and −82. Also, a 11 bp repeat beginning at positions −126 and −99 and a 15 bp repeat starting at positions −234 and −171 (with a single base difference) were found.

The nucleotide sequence reported here showed complete identity in the overlapping regions (from −126 to 210) with the sequence of an α1(II) genomic clone published by Sandell *et al.* (1984). The deduced amino acid sequence of the triple-helical and C-telopeptide region based on nucleotide sequences from −369 to −283 and from −237 to −1 is also identical with the partial amino acid sequence of chicken type II collagen determined by W. T. Butler (personal communication) at amino acid positions 919–947 and 968–1041.

Comparison of nucleotide and amino acid sequences of chick pro-α1(II) and other procollagens

By using published nucleotide and deduced amino acid sequences for human α1(II) (Strom & Upholt, 1984) and chick α1(I), α2(I) (Fuller & Boedtker, 1981) and α1(III) (Yamada *et al.*, 1983) procollagens, as well as the composite α1(II) sequence from pCgII-S01 and pCgII-S13 (Fig. 3), comparative analyses were performed. During sequence alignment spaces were inserted at certain positions to increase the number of amino acid matches.

An interspecies comparison of chick α1(II) procollagen with the available portion of human α1(II) procollagen (Strom & Upholt, 1984) indicated a high degree of overall nucleotide (84%; 129/153 bases) and amino acid (84%; 48/51 amino acids) sequence homology. Most of the nucleotide substitutions appeared in 'silent' sites (17/24), indicating a strong selective pressure on the conservation of the *N*-terminal portions of the C-propeptide.

In order to determine the location of the most conserved regions, which may perform a structural or functional role common for the different collagens, all four chicken procollagens were compared and heterogeneities calculated at intervals of seven amino acid residues. Statistical evaluation of the amino acid heterogeneity is depicted graphically in Fig. 4. Heterogeneity was the highest in the telopeptide region and the *N*-terminus of the C-propeptide. Within this overall variable region there is a rather conserved area corresponding to the C-proteinase-cleavage site.

The helical region was not uniformly conserved. Regions having low degrees of heterogeneity were found periodically at distances of 20–25 amino acids. The position of the most conserved regions correlated with the location of hydroxyproline, hydroxylysine, lysine and cysteine residues (W. T. Butler, personal communication). This indicated a strong selective pressure for conservation of these residues, which are involved in inter- and intra-chain interactions. Conserved regions may serve functions common to all types of collagens. The high variability of the telopeptide suggested that it either may perform a type-specific function; or that the same function shared by all collagens can be accomplished by a number of different structural forms.

In order to determine the degree of divergence between the different chicken procollagen genes, nucleotide sequences were compared in pairs and the corrected divergence for both 'silent' and replacement sites was determined by the method of Perler *et al.* (1980) as modified by Bernard *et al.* (1983*b*). As shown in Table 1, the corrected divergence of replacement sites is significantly less between α1(II) and α1(I) procollagens than between α1(I) and α2(I) or between α1(II) and α1(III) procollagens ($P < 0.05$). Similarly, the corrected divergence of 'silent' sites is significantly less between α1(II) and α1(I) procollagens than between α1(II) and α2(I) or α1(I) and α2(I) procollagens. On the basis of these calculations it was concluded that genes coding for α1(II) and α1(I) procollagens had been separated from each other later during evolution than those coding for α1(I) and α2(I) procollagens. This is in agreement with earlier observations based on amino acid sequence analysis (Butler *et al.*, 1977; Bornstein & Traub, 1979).

Assuming that the number of replacement-site mutations is proportional to the evolutionary time (Perler *et al.*, 1980), one can determine divergence time of two genes, provided that appropriate reference points can be found. Bernard *et al.* (1983*b*) published that the corrected divergence of replacement sites calculated by comparing human, mouse and chicken type I procollagen sequences increased proportionally with the time of mammalian–avian divergence and mammalian radiation based on fossil data. They reported 27% corrected divergence for chicken α1(I) and α2(I) C-propeptide coding sequences and concluded that the two genes started to diverge $950 \times 10^6 \pm 120 \times 10^6$ years ago (Bernard *et al.*, 1983*b*). A somewhat lower value, 730×10^6 years, was obtained by Mathews (1980), this being based on amino acid sequence comparison of chains from different species. We found $18 \pm 1.9\%$ corrected nucleotide divergence for replacement sites of α1(II) and α1(I) procollagens, which were the nearest relatives in the comparison. On the basis of the time curve determined by Bernard *et al.* (1983*b*), we can assume that the α1(I) and α1(II) chain divergence began $630 \times 10^6 \pm 70 \times 10^6$ years ago. This period is

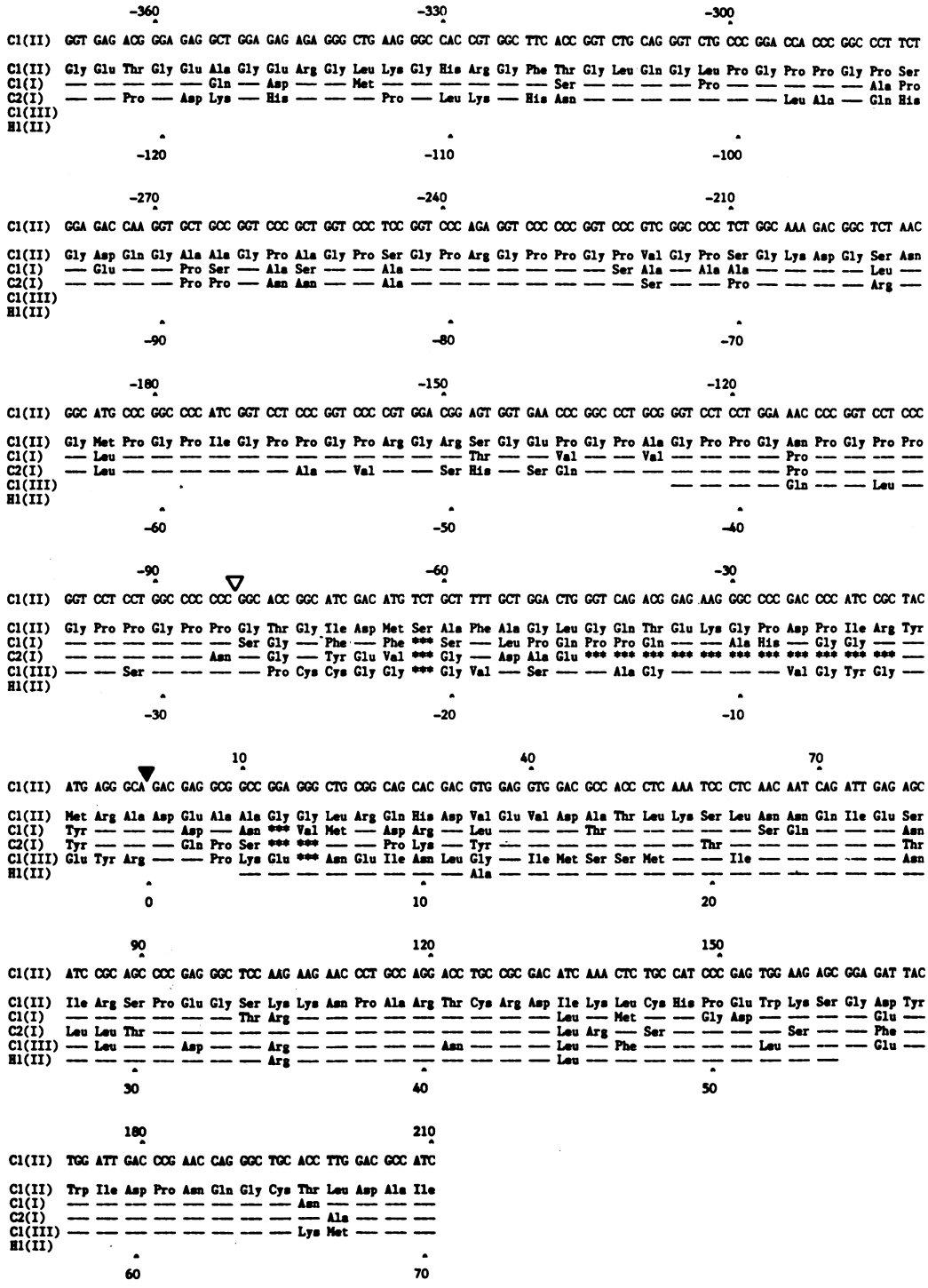

Fig. 3. *Nucleotide and amino acid sequences of chicken type II procollagen*

The nucleotide sequence coding for part of the helical domain, the entire telopeptide and a portion of the C-propeptide of chicken type II procollagen [C1(II)] is aligned with the corresponding amino acid sequences of human type II procollagen [H1(II)] (Strom & Upholt, 1984) and other chicken procollagens [C1(I), C2(I) and C1(III)] (Fuller & Boedtker, 1981; Yamada *et al.*, 1983). The sequences are numbered from the C-proteinase-cleavage site (▼). ▽, Border between the helical domain and the telopeptide. ***, Space inserted during alignment in order to increase the number of matches; —, the same amino acid as in C1(II).

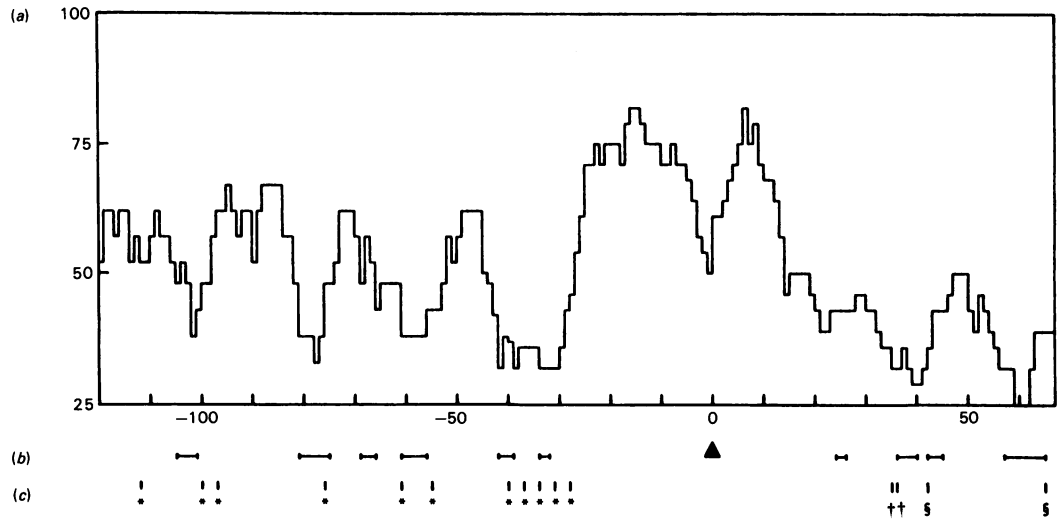

Fig. 4. *Graphical representation of amino acid sequence homology of α1(II), α2(I) and α1(III) procollagens aligned as in Fig. 3*
(*a*) A graph of amino acid heterogeneity. Values represent the number of different amino acids as a percentage of all amino acids present in intervals of seven amino acid residues. Note that between positions −120 and −42 three sequences were compared, and therefore the theoretical minimum value is 33%, whereas in the rest of the graph, where four sequences were compared, the minimum is 25%. Terminal amino acids that do not have three amino acids to each side are not included in the graph. Spaces inserted for better alignment (see Fig. 3) were considered as differences. (*b*) Bars (|——|) represent amino acid stretches longer than three residues identical in all sequences. (*c*) Positions of hydroxyproline or hydroxylysine (*), lysine (†) and cysteine (§) residues in type II procollagen (W. T. Butler, personal communication). ▲, C-proteinase-cleavage site.

Table 1. *Comparison of amino acid and nucleotide differences between chick procollagens*
Nucleotide and amino acid sequences of the C-propeptide, the C-telopeptide and 96 amino acids of the helical domain of α1(I) and α1(I) (Fuller & Boedtker, 1981), α1(II) (present work; Ninomiya *et al.*, 1984*b*) and α1(III) (Yamada *et al.*, 1983) procollagens were compared. Spaces inserted to increase the numbers of matches at the amino acid level were not included in the calculation. Nucleotide change values are variance-weighted means ± sampling error (Bernard *et al.*, 1983*b*).

	Amino acid replacements (% difference)	Nucleotide changes (% corrected divergence)	
		Replacement sites	'Silent' sites
α1(II)–α1(I)	29	18 ± 1.9	62 ± 6.7
α1(II)–α2(I)	33	24 ± 1.9	117 ± 14
α1(II)–α1(III)	38	29 ± 2.4	112 ± 25
α1(I)–α2(I)	38	27 ± 2.1	117 ± 13

identical with that estimated for the appearance of the first Metazoa (Cloud, 1968). The idea that this divergence may have been one prerequisite for multicellularity is intriguing. It must be noted, however, that the calculation was based on the assumption that the divergence rate for replacement sites did not change during evolution. One cannot rule out the possibility that the mutation rate was higher at the time of appearance of multicellular organisms. A comparison of α1(II) sequence divergence between other species would be required for a more precise determination of the exact time of divergence of the α1(II) from other types of collagen genes.

We are grateful to Dr. B. R. Olsen for providing us with a manuscript ahead of publication and to Dr. William T. Butler for unpublished amino acid sequence information on chick type II collagen. This work was supported by Grant HD09174 from the National Institute of Child Health and Human Development. This paper is Scientific Contribution no. 1083 from the Storrs Agricultural Experiment Station (University of Connecticut, Storrs, CT 06268, U.S.A.).

References

Berkner, K. L. & Folk, W. R. (1977). *J. Biol. Chem.* **252**, 3176–3184

Bernard, M. P., Myers, J. C., Chu, M.-L., Ramirez, F., Eikenberry, E. F. & Prockop, D. J. (1983*a*) *Biochemistry* **22**, 1139–1145

Bernard, M. P., Chu, M.-L., Myers, J. C., Ramirez, F., Eikenberry, E. F. & Prockop, D. J. (1983*b*) *Biochemistry* **22**, 5213–5223

Biggin, M. D., Gibson, T. J. & Hong, G. F. (1983) *Proc. Natl. Acad. Sci. U.S.A.* **80**, 3963–3965

Bornstein, P. & Sage, H. (1980) *Annu. Rev. Biochem.* **49**, 957–1003

Bornstein, P. & Traub, W. (1979) in *Proteins*, vol. 4 (Neurath, H. & Hill, R., eds.), pp. 411–632, Academic Press, New York and London

Boyd, C. D., Tolstoshev, P., Schafer, M. P., Trapnell, B. C., Coon, H. C., Kretschmer, P. J., Nienhuis, A. W. & Crystal, R. G. (1980) *J. Biol. Chem.* **255**, 3212–3220

Butler, W. T., Finch, J. E., Jr. & Miller, E. J. (1977) *Biochemistry* **16**, 4981–4990

Clewell, D. B. & Helinski, D. R. (1972) *J. Bacteriol.* **110**, 1135–1141

Cloud, P. E. (1968) *Science* **160**, 729–736

Deák, F., Argraves, W. S., Kiss, I. & Goetinck, P. F. (1984) *Fed. Proc. Fed. Am. Soc. Exp. Biol.* **43**, 1983

Fessler, J. H. & Fessler, L. I. (1978) *Annu. Rev. Biochem.* **47**, 129–162

Fuchs, R. & Blakesley, R. (1983) *Methods Enzymol.* **100**, 3–38

Fuller, F. & Boedtker, H. (1981) *Biochemistry* **20**, 996–1006

Hanahan, D. (1983) *J. Mol. Biol.* **166**, 557–580

Hanahan, D. & Meselson, M. (1983) *Methods Enzymol.* **100**, 333–342

Helfman, D. M., Feramisco, J. R., Fiddes, J. C., Thomas, G. P. & Hughes, S. H. (1983) *Proc. Natl. Acad. Sci. U.S.A.* **80**, 31–35

Lukens, L. N., Frischauf, A. M., Pawlowski, P. J., Bierley, G. T. & Lehrach, H. (1983) *Nucleic Acids Res.* **11**, 6021–6039

Mathews, M. B. (1980) in *Biology of Collagen* (Viidik, A. & Vuust, J., eds.), pp. 193–209, Academic Press, New York

Maxam, A. & Gilbert, W. (1980) *Methods Enzymol.* **65**, 99–560

Monson, J. M., Friedman, J. & McCarthy, B. J. (1982) *Mol. Cell. Biol.* **2**, 1362–1374

Ninomiya, Y., Showalter, A. M. & Olsen, B. R. (1984*a*) *Symp. Soc. Dev. Biol.* **42**, 255–275

Ninomiya, Y., Showalter, A. M., van der Rest, M., Seidah, N. G., Chrétien, M. & Olsen, B. R. (1984*b*) *Biochemistry* **23**, 617–624

Perler, F., Efstratiadis, A., Lomedico, P., Gilbert, W., Kolodner, R. & Dodgson, J. (1980) *Cell (Cambridge, Mass.)* **20**, 555–566

Sandell, L. J., Yamada, Y., Dorfman, A. & Upholt, W. B. (1983) *J. Biol. Chem.* **258**, 11617–11621

Sandell, L. J., Prentice, H. L., Kravis, D. & Upholt, W. B. (1984) *J. Biol. Chem.* **259**, 7826–7834

Strom, C. M. & Upholt, W. B. (1984) *Nucleic Acids Res.* **12**, 1025–1038

Tanaka, T. & Weisblum, B. (1975) *J. Bacteriol.* **121**, 354–362

Thomas, P. S. (1980) *Proc. Natl. Acad. Sci. U.S.A.* **77**, 5201–5205

Vuorio, E., Sandell, L., Kravis, D., Sheffield, V. C., Vuorio, T., Dorfman, A. & Upholt, W. B. (1982) *Nucleic Acids Res.* **10**, 1175–1192

Yamada, T., Kuhn, K. & de Crombrugghe, B. (1983) *Nucleic Acids Res.* **11**, 2733–2744